

# A\* Orthogonal Matching Pursuit: Best-First Search for Compressed Sensing Signal Recovery

Nazim Burak Karahanoglu<sup>a,b,\*</sup>, Hakan Erdogan<sup>a</sup>

<sup>a</sup>*Department of Electronics Engineering, Sabanci University, Istanbul 34956, Turkey*

<sup>b</sup>*Information Technologies Institute, TUBITAK-BILGEM, Kocaeli 41470, Turkey*

---

## Abstract

Compressed sensing is a developing field aiming at reconstruction of sparse signals acquired in reduced dimensions, which make the recovery process under-determined. The required solution is the one with minimum  $\ell_0$  norm due to sparsity, however it is not practical to solve the  $\ell_0$  minimization problem. Commonly used techniques include  $\ell_1$  minimization, such as Basis Pursuit (BP) and greedy pursuit algorithms such as Orthogonal Matching Pursuit (OMP) and Subspace Pursuit (SP). This manuscript proposes a novel semi-greedy recovery approach, namely A\* Orthogonal Matching Pursuit (A\*OMP). A\*OMP performs A\* search to look for the sparsest solution on a tree whose paths grow similar to the Orthogonal Matching Pursuit (OMP) algorithm. Paths on the tree are evaluated according to a cost function, which should compensate for different path lengths. For this purpose, three different auxiliary structures are defined, including novel dynamic ones. A\*OMP also incorporates pruning techniques which enable practical applications of the algorithm. Moreover, the adjustable search parameters provide means for a complexity-accuracy trade-off. We demonstrate the reconstruction ability of the proposed scheme on both synthetically generated data and images using Gaussian and Bernoulli observation matrices, where A\*OMP yields less reconstruction error and higher exact recovery frequency than BP, OMP and SP. Results also indicate that novel dynamic cost functions provide improved results as compared to a conventional choice.

**Keywords:** compressed sensing, sparse signal reconstruction, orthogonal matching pursuit, best-first search, auxiliary functions for A\* search

---

## 1. Introduction

Compressed sensing (CS) deals with the acquisition of the sparse signals, i.e. signals with only a few nonzero coefficients, in reduced dimensions. As a natural consequence of this, the signal has to be reconstructed back to its full dimension using the observation in reduced dimensions. CS is based on the following question: Can a reduced number of observations (less than Shannon-Nyquist rate) contain enough information for exact reconstruction of sparse signals? One might argue that this seems quite unnatural, however a number of articles in CS literature, i.e. [1], [2] and [3], state that it is indeed possible under certain assumptions.

Exact solution of the CS reconstruction problem requires minimization of the  $\ell_0$  norm, i.e. the number of nonzero coefficients, which is impractical. One of the solutions that can be found in the literature is the convex relaxation which replaces  $\ell_0$  minimization problem with an  $\ell_1$  minimization, such as Basis Pursuit [4]. Another family of algorithms, so called greedy pursuit algorithms, Orthogonal Matching Pursuit (OMP) [5], Subspace Pursuit (SP) [6], Iterative Hard Thresholding (IHT) [7, 8] etc. provide greedy and find approximate solutions by solving a stagewise constrained residue minimization problem.

This manuscript proposes a new semi-greedy CS reconstruction approach that incorporates the A\* Search [9, 10, 11, 12, 13], a best-first search technique that is frequently used in path finding, graph traversal and speech recognition.

---

\*Corresponding author

Email addresses: karahanoglu@sabanciuniv.edu (Nazim Burak Karahanoglu), herdogan@sabanciuniv.edu (Hakan Erdogan)

This new method, which we call A\*OMP, proposes an A\* search that employs the OMP algorithm to expand the most promising path of the search tree at each iteration. By utilizing best-first search, multiple paths can be evaluated during the search, which promises improvements over the single path structures of algorithms such as MP or OMP. This combination of A\* search and OMP is not straightforward: It requires appropriately defined cost models which enable A\* to perform stage-wise residue minimization in an intelligent manner, and effective pruning techniques which make the algorithm tractable in practice. As for the cost model, which should make comparison of paths with different lengths possible, we introduce two novel dynamic structures, which better comply with our needs, in addition to the trivial additive one. Pruning capability is provided via a number of strategies which, together with the cost model parameters, enable a complexity-accuracy trade-off. The effectiveness of the proposed pruning techniques and the dynamic cost models is demonstrated via provided reconstruction examples. This reconstruction experiments, including different nonzero coefficient distributions, Gaussian and Bernoulli type random observation matrices, noise contaminated measurements and images, demonstrate that utilization of best-first search is able to improve the reconstruction accuracy. A preliminary version of this work has been presented in [14].

A number of tree-search based methods have appeared in CS literature. These methods are, however, fundamentally different than A\*OMP as they do not follow the best-first search principle. The tree-search based OMP (TB-OMP), [15], employs a tree-search that opens  $L$  children per each node at a level. A rather flexible version of this is the flexible tree-search based OMP (FTB-OMP) [16], where the branching factor  $L$  is decreased at each level. Another straightforward tree-search also appears in Fast Bayesian Matching Pursuit [17], which opens all children of the nodes at a level, and retains the best  $D$  wrt. their posterior probabilities. These methods incorporate rather simple and non-sophisticated tree-search techniques in comparison to A\*OMP. They employ neither cost models to compensate for different path lengths, nor mechanisms to select the most promising path on the fly, but expand all nodes at a level. They do not also possess effective pruning techniques, except FTB-OMP pruning the children of a node wrt. their correlations to the best one, and FBMP keeping  $D$  nodes at a level. The randomized OMP (RandOMP) algorithm [18] yields an estimate of the minimum mean-squared error (MMSE) solution by averaging multiple sparse representations which are obtained by running a randomized version of OMP several times. Though RandOMP involves multiple sparse representations, it incorporates no explicit tree-search.

To avoid some possible misunderstanding, we would like to note that the tree search concept in A\*OMP is completely general to all sparse signals. A\*OMP aims to find a closer result to the true  $\ell_0$  solution, thus the objective is to improve reconstruction quality not to decrease computational complexity to find a greedy solution, such as in list decoding [19]. Furthermore, A\*OMP is neither specific for tree-sparse signals nor does it make use of a tree-structured over-complete basis as for the tree-based OMP algorithm [20]. The algorithm is not specific for structured sparse signals as well.

The rest of this manuscript is organized as follows: CS reconstruction problem and some major algorithms are introduced briefly in sections 2. A\* search is discussed in section 3. Section 4 is devoted to the A\*OMP algorithm and the novel cost functions. We demonstrate the reconstruction performance of A\*OMP in comparison to Basis Pursuit (BP) [4], Subspace Pursuit (SP) [6] and OMP [5] in Section 5, before concluding the manuscript with a short summary.

## 2. Compressed Sensing

### 2.1. Problem Definition

Compressed Sensing acquisition of a  $K$ -sparse signal  $\mathbf{x}$ , i.e. having only  $K$  nonzero entries, is obtained via the observation matrix, or dictionary,  $\Phi$ :

$$\mathbf{y} = \Phi \mathbf{x} \quad (1)$$

where  $\mathbf{x} \in \mathbb{R}^N$ ,  $\Phi \in \mathbb{R}^{M \times N}$ ,  $\mathbf{y} \in \mathbb{R}^M$  and  $K < M < N$ . As  $M < N$ , solving for  $\mathbf{x}$  directly from (1) is ill-posed. CS exploits sparsity of  $\mathbf{x}$  to formulate the reconstruction problem alternatively as

$$\mathbf{x} = \arg \min \|\mathbf{x}\|_0 \quad s.t. \quad \mathbf{y} = \Phi \mathbf{x}. \quad (2)$$

where  $\|\cdot\|_0$  denotes the  $\ell_0$  norm, which is the number of nonzero coefficients of a signal. Solving (2) directly is not feasible as it requires an exhaustive combinatorial search [1, 21]. Consequently, a variety of strategies have emerged to find approximate solutions to (2).

## 2.2. Theoretical Guarantees - The Restricted Isometry Property

An important means for obtaining theoretical guarantees in CS recovery problem is the restricted isometry property (RIP) [22, 3, 23]: A matrix  $\Phi$  is said to satisfy the  $K$ -RIP if there exists a restricted isometry constant  $\delta_K$ ,  $0 < \delta_K < 1$  such that

$$(1 - \delta_K)\|\mathbf{x}\|_2^2 \leq \|\Phi\mathbf{x}\|_2^2 \leq (1 + \delta_K)\|\mathbf{x}\|_2^2, \forall \mathbf{x} : \|\mathbf{x}\|_0 \leq K. \quad (3)$$

A matrix satisfying the RIP acts almost like an orthonormal system for sparse linear combinations of its columns [22], making reconstruction of sparse signals from lower dimensional observations possible.

Analysis in [3, 24] state that matrices with i.i.d. Gaussian or Bernoulli entries and matrices randomly selected from discrete Fourier transform satisfy the RIP with high probabilities, when they satisfy some specific conditions on  $K$  based on  $M$  and  $N$ . Therefore, such random observation matrices can provide compact representations of sparse signals.

## 2.3. Major CS Reconstruction Algorithms

Following [25], CS recovery approaches can be categorized as greedy pursuit algorithms, convex relaxation, Bayesian framework, nonconvex optimization and brute force methods. In this work, we are interested in the first two of these.

### 2.3.1. Convex Relaxation

$\ell_1$  or convex relaxation algorithms rely on the relaxation of the  $\ell_0$  norm minimization in (2) by an  $\ell_1$  norm, which first appeared in Basis Pursuit [4]. In this context, (2) is rewritten as

$$\mathbf{x} = \arg \min \|\mathbf{x}\|_1 \quad s.t. \quad \mathbf{y} = \Phi\mathbf{x}, \quad (4)$$

which can be solved via computationally tractable convex optimization methods, such as pivoting, linear programming and gradient methods [25]. Extensive analysis of RIP conditions for  $\ell_1$  relaxation can be found in [22, 3, 23, 26].

### 2.3.2. Greedy Pursuits

Historically, Matching Pursuit (MP) [27] is the first greedy pursuit. MP expands the support of  $\mathbf{x}$  by the dictionary atom which has the highest inner-product with the residue at each iteration. Major drawback of MP is that it does not take into account the non-orthogonality of the dictionary, which results in suboptimal choices of the nonzero coefficients.

The non-orthogonality of dictionary atoms is taken into account by the Orthogonal Matching Pursuit (OMP) [5], which performs orthogonal projection of the residue onto the selected dictionary atoms after each iteration. Ensuring orthogonality of the residue to the selected support enhances the reconstruction. As expansion of paths in A\*OMP is very similar to OMP, we devote some space to a short overview of this method.

Let's first define the notation: Let  $\mathbf{v}_n \in \mathbb{R}^M$ ,  $n = 1, 2, \dots, N$  be the dictionary atoms, i.e. columns of the dictionary  $\Phi$ .  $\mathbf{r}^l$  denotes the residue after the  $l$ th iteration.  $\mathbf{S}$  and  $\mathbf{c}$  denote the matrix (or, exchangeably in context, the set) of atoms selected from  $\Phi$  for representing  $\mathbf{y}$  and the vector of corresponding coefficients respectively.

OMP is initialized as  $\mathbf{r}^0 = \mathbf{y}$ ,  $\mathbf{S} = \{\}$  and  $\mathbf{c} = \mathbf{0}$ . At iteration  $l$ , OMP appends  $\mathbf{S}$  the dictionary atom that best matches  $\mathbf{r}^{l-1}$

$$\begin{aligned} \mathbf{s} &= \arg \max_{\mathbf{v}_n \in \Phi \setminus \mathbf{S}} \langle \mathbf{r}^{l-1}, \mathbf{v}_n \rangle, \\ \mathbf{S} &= \mathbf{S} \cup \mathbf{s}. \end{aligned} \quad (5)$$

The coefficients are computed by the orthogonal projection

$$\mathbf{c} = \arg \min_{\tilde{\mathbf{c}} \in \mathbb{R}^l} \|\mathbf{y} - \mathbf{S}\tilde{\mathbf{c}}\|_2. \quad (6)$$

At the end of each iteration, the residue is updated as

$$\mathbf{r}^l = \mathbf{y} - \mathbf{S}\mathbf{c}. \quad (7)$$

After termination,  $\mathbf{S}$  and  $\mathbf{c}$  contain the support and the corresponding nonzero entries of  $\mathbf{x}$ , respectively. OMP may employ different termination criterion. In this work, we fix the number of iterations as  $K$ . Alternatively, iterations can be carried on until the residue falls below a threshold.

A detailed analysis of OMP is provided in [28] which states a lower-bound on the number of observations for exact recovery. The guarantees for OMP, however, were shown to be non-uniform, i.e. they hold only for each fixed sparse signal, but not for all [29]. It was shown in [30] that for natural random matrices it is not possible to obtain uniform guarantees for OMP.

Recently, more sophisticated pursuit methods, which select multiple columns per iteration, have appeared. For example, Stagewise OMP (StOMP) [31] selects in each step all columns whose inner-products with the residue is higher than an adaptive threshold depending on the  $\ell_2$  norm of the residue. Alternatively, regularized OMP (ROMP) [29] groups inner-products with similar magnitudes into sets at each iteration and selects the set with maximum energy. Via this regularization, ROMP provides RIP-based uniform guarantees. Compressive Sampling Matching Pursuit (CoSaMP) [32] and Subspace Pursuit (SP) [6] combine selection of multiple columns per iteration with a pruning step. At each iteration, these first expand the selected support by addition of new atoms, and then prune it to retain only the best  $K$  atoms. Both CoSaMP and SP are provided with optimal performance guarantees based on RIP.

Iterative hard thresholding (IHT) [7, 8] employs an iterative gradient search that first updates the sparse estimate in the direction of the gradient of the residue wrt. the dictionary and then prunes the solution by either thresholding or keeping only the  $K$  largest entries. IHT is equipped with RIP based guarantees similar to CoSaMP and SP [8]. A recent IHT variant, Nesterov iterative hard thresholding (NIHT) [33] employs Nesterov's proximal gradient [34] to update the sparse representation. NIHT provides no a priori performance guarantee, but still an online performance guarantee.

### 3. A\* Search

A\* search [9, 10, 11, 12, 13] is an iterative tree-search algorithm. In our problem, the A\* search tree is iteratively built up by nodes which represent the dictionary atoms. Each path from the root to a leaf node denotes a subset of dictionary atoms which is a candidate support for  $\mathbf{x}$ . A path is called *complete* if it has  $K$  nodes, and *partial* if it is shorter. A\* search tree is initialized with all possible single-node paths. At each iteration, the most promising path is chosen and all of its children are added to the search tree. Search is terminated when the most promising path is found to be complete.

In our scope, A\* search looks for the complete path  $\mathbf{p}^K$  which minimizes some evaluation function  $g(\mathbf{p}^K)$ . As tree paths typically have different lengths, these cannot be compared via an evaluation function which depends on the number of nodes on the path. In order to deal with different path lengths, A\* search employs an auxiliary function [11]. For a path  $\mathbf{p}^l$  of length  $l \leq K$ , the auxiliary function  $d(\mathbf{p}^l)$  is defined such that  $d(\mathbf{p}^K) = 0$  and

$$d(\mathbf{p}^l) \geq g(\mathbf{p}^l) - g(\mathbf{p}^l \cup \mathbf{z}^{K-l}), \quad \forall \mathbf{z}^{K-l}, \quad (8)$$

where  $\mathbf{z}^{K-l}$  is a sequence of  $K - l$  nodes and  $\cup$  denotes concatenation. With this definition,  $d(\mathbf{p}^l)$  is larger than or equal to the decrement in the evaluation function that any complete extension of the path  $\mathbf{p}^l$  could yield.

Now, we define the cost function as

$$F(\mathbf{p}^l) = g(\mathbf{p}^l) - d(\mathbf{p}^l). \quad (9)$$

Let's consider a complete path  $\mathbf{p}^K$  and a partial path  $\tilde{\mathbf{p}}^l$  of length  $l < K$ . Combining (8) and (9), if  $F(\mathbf{p}^K) \leq F(\tilde{\mathbf{p}}^l)$ , then  $g(\mathbf{p}^K) \leq g(\tilde{\mathbf{p}}^l \cup \mathbf{z}^{K-l})$  for all  $\mathbf{z}^{K-l}$ , which states that  $\mathbf{p}^K$  is better than all possible extensions of  $\tilde{\mathbf{p}}^l$ . Hence, it is safe to use the cost function  $F(\cdot)$  for selecting the most promising path. Note that, satisfying (8) may either be impossible or unpractical in practice. This issue is discussed when different A\*OMP cost models are introduced in Section 4.3.

### 4. Sparse Signal Reconstruction using A\* Search

A\*OMP casts the sparse recovery problem into a search for the correct support of the  $K$ -sparse  $\mathbf{x}$  among a number of dynamically evolving candidate subsets. These candidate subsets are stored as paths from the root node to leaf nodes of a search tree, where each node represents an atom in  $\Phi$ . The search tree is built up and evaluated iteratively

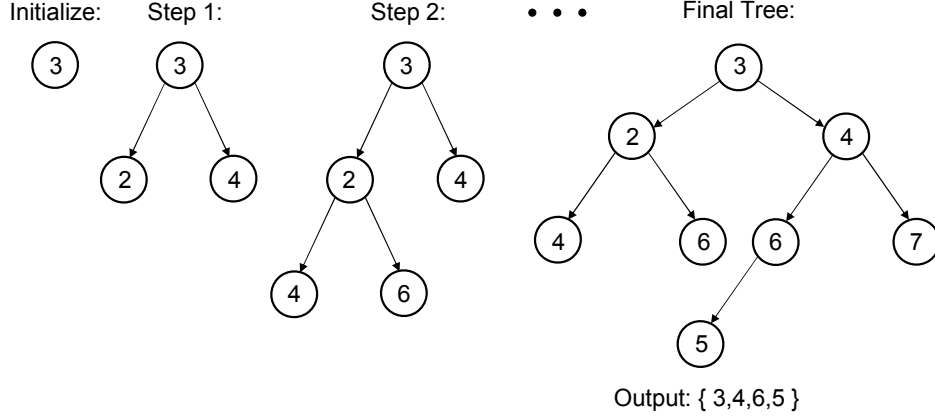


Figure 1: Evaluation of the search tree during A\*OMP algorithm

by A\* search. The search starts with candidate subsets of single elements. At each iteration, new dictionary atoms are appended to the most promising path, which is selected to minimize some cost function based on the residue. In this way, A\*OMP performs a multi-path search for the best one among all possible  $K$ -element subsets of  $\Phi$ . Though the A\*OMP search tree actually restricts the search to a set of iteratively built candidate subsets, it is general with the capability of representing all possible  $K$ -element subsets of  $\Phi$ . Fig. 1 illustrates evaluation of a sample search tree throughout the search.

Incorporation of a multi-path search strategy is motivated by the expectation that it would improve reconstruction especially where a single-path algorithm such as OMP fails because of the linear dependency of dictionary atoms. In cases where computation of a single path yields a wrong representation, the correct one will mostly be in the set of candidate representations. By a properly configured multi-path search, i.e. by proper selection of the cost model as discussed below, this correct path may be distinguished among the candidates. In other words, a multi-path strategy may reduce the error especially when too few measurements are provided.

For the rest of this work, we differentiate between the paths in the search tree with subscripts. The superscripts represent either the length of the path, or the position of the node in the path.  $s_i^l$  represents the selected atom at the  $l$ 'th node on path  $S_i$  and  $c_i^l$  the corresponding coefficient. Similarly  $r_i$  is the residue of path  $i$ .  $S_i$  and  $c_i$  denote the matrix of atoms selected for path  $i$  and the vector of corresponding coefficients, respectively. Note that  $S_i$  and  $s_i^l$  are the mathematical equivalents of the corresponding path and node, respectively. In the rest of this work, we slightly abuse this notation and use  $s_i^l$  and  $S_i$  also to represent the corresponding node and path.

We discuss utilization of tree search for A\*OMP in three main steps: initialization of the search tree, selecting the best path and expansion of the selected partial path.

#### 4.1. Initialization of the Search Tree

A\* search originally initializes the search tree by all possible paths with length 1. This corresponds to  $N$  different initial subsets, which is not practical in most cases as  $N$  is usually large. In fact, only  $K \ll N$  dictionary atoms are relevant to  $\mathbf{y}$ . Moreover, each iteration adds the tree multiple children of a selected partial path (Section 4.2). Hence, the search might be started with less paths. As a consequence, we limit the initial search tree to the  $I \ll K$  subsets, each of which contains one of the  $I$  atoms having the highest absolute inner-product with  $\mathbf{y}$ . Note that another possibility would be selecting the atoms whose inner-products with  $\mathbf{y}$  are greater than a certain threshold.

#### 4.2. Expanding the Selected Partial Path

In typical A\* search, all children of the most promising partial path are added to the search tree at each iteration. In practice, this results in too many search paths because of the high number of possible children: To illustrate, let the length of the selected partial path be  $l$ . This path has  $N - l \approx N$  children since  $l < K \ll N$ . Hence, each iteration considers approximately  $N$  new paths and the upper bound on the number of paths involved overall in the search is obtained as  $N^K$ , given  $K \ll N$ . To limit these, we employ three pruning strategies:

#### 4.2.1. Extensions per Path Pruning

For our purposes, order of nodes along a path is unimportant. At each step, we require only to add one of the  $K$  correct atoms to the representation, and not a specific one of them. Therefore, considering only a few children of a selected partial path becomes a reasonable sacrifice. At each A\*OMP iteration, we expand the search tree only by the  $B$  children which have the highest absolute inner-product with the residue to the selected path. Note that another reasonable choice would be considering only the children whose inner-products with the residue are higher than a threshold.

Extensions per Path Pruning decreases the upper bound on the number of paths from  $N^K$  to  $B^K$ . Starting the search with  $I$  initial paths, this bound becomes  $I * B^{(K-1)}$ . Practically,  $I$  and  $B$  are chosen much smaller than  $N$ , decreasing the paths involved in the search drastically.

#### 4.2.2. Tree Size Pruning

Despite extensions per path are limited to  $B$ , adding new paths at each iteration still increases required memory, as the corresponding residues are also necessary. To reduce memory requirements, we adopt the ‘‘beam search’’ strategy and we limit the maximum number of paths in the tree by the beam width  $P$ . When this limit is exceeded, the worst paths, i.e. the ones with maximum cost, are removed from the tree till  $P$  paths remain.

Fig. 2 illustrates the Extensions per Path and Tree Size Pruning rules where  $P = 4$  and  $B = 3$ . Fig. 2a depicts a search tree with four paths at the beginning of an iteration. The cost of each path is indicated with  $C_i$ . Path 4, which has the minimum cost is selected as the best path. Let the best  $B$  children of path 4 be nodes 2, 8 and 9, ordered with descending correlation to the residue. In Fig. 2b, the best child 2 is directly appended to Path 4, without increasing the number of paths. Fig. 2c depicts addition of the second child 8, after which there appear five paths on the tree. As tree size is limited to  $P = 4$ , path 2, which has the maximum cost, is removed. Finally, we consider node 9 in Fig. 2d. The resultant path has higher cost than the other four paths. Hence, it is not added to the tree.

#### 4.2.3. Equivalent Path Pruning

Neglecting insertion of equivalent paths to the tree is also important to improve the search performance. For this purpose, we define a path equivalency notion that also covers paths with different lengths: Let  $S_1^{l_1}$  and  $S_2^{l_2}$  be two paths of lengths  $l_1$  and  $l_2$ , respectively, where  $l_1 \geq l_2$ . Let’s define  $S_{p,1}^{l_2}$  as the partial path that consists of the first  $l_2$  nodes of  $S_1^{l_1}$ , i.e.  $S_{p,1}^{l_2} = s_1^1, s_1^2, \dots, s_1^{l_2}$ .  $S_1^{l_1}$  and  $S_2^{l_2}$  are equivalent if and only if  $S_{p,1}^{l_2}$  and  $S_2^{l_2}$  share the same set of nodes. In this case, orthogonality of the residue to the selected support, ensures that  $S_{p,1}^{l_2}$  and  $S_2^{l_2}$  are equivalent. Consequently, insertion of  $S_2^{l_2}$  into the tree is unnecessary, as  $S_{p,1}^{l_2}$  has already been expanded in previous iterations.

Fig. 3 illustrates the path equivalency. Path 2 and the first three nodes of Path 1 share the same set of nodes, which makes Path 1 and Path 2 equivalent. Note that orthogonal projection ensures node 5 will be among the best children of path 2. On the contrary, Path 1 and Path 3 are not equivalent as the first three nodes of Path 1 and Path 3 are different. There exists no guarantee that node 7 will be among the best children of Path 3.

Let’s now summarize extension of a selected partial path with these three pruning rules: First, the best  $B$  children of the selected partial path  $S$  are chosen as the dictionary atoms having highest inner-product with the residue. We obtain  $B$  new candidate paths by appending  $S$  one of these  $B$  children. We apply Equivalent Path Pruning by eliminating candidates which are equivalent to already visited paths. For each remaining candidate, we first compute the residue via orthogonal projection of  $\mathbf{y}$  onto  $S$ , and then the cost as discussed below. We remove  $S$  from the tree and add the candidate paths. Finally, we prune the tree if number of paths exceeds  $P$ .

#### 4.3. Selection of the Most Promising Path

A natural criterion for choosing the most promising path is the minimum residual error. Consequently, for a path  $\mathbf{S}^l$  of length  $l$ , the evaluation function can be written as

$$g(\mathbf{S}^l) = \|\mathbf{r}^l\|_2 = \left\| \mathbf{y} - \sum_{j=1}^l c^j \mathbf{s}^j \right\|_2. \quad (10)$$

where  $\mathbf{s}^j$  and  $c^j$  denote the selected atom at stage  $j$  and the coefficient obtained after orthogonal projection of the residue onto the set of selected atoms, respectively.

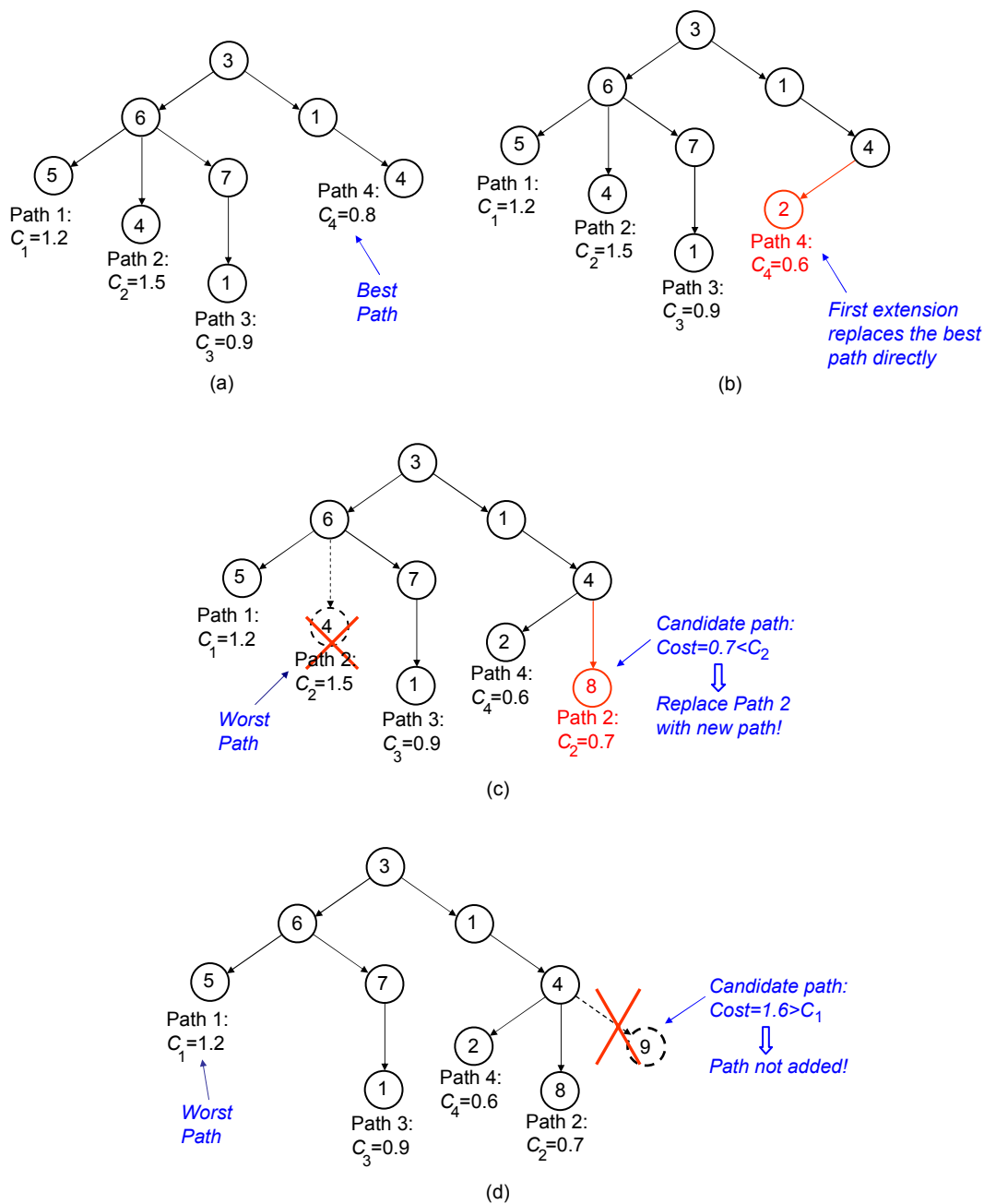


Figure 2: Evaluation of the search tree during a single iteration of the A\*OMP algorithm

As discussed in Section 3, A\* search employs an auxiliary function to compensate for different path lengths. The auxiliary function is important for comparing the multiple paths in the search tree. By proper evaluation of these paths, though any single one of them is limited to the RIP condition of OMP algorithm alone, A\*OMP can relax the RIP condition, increasing the probability of finding a final path that is not altered by the linear dependency of the atoms in the dictionary. Ideally, the auxiliary function should mimic the decay of the residue along a path, which is impossible in practice. Below, we suggest three different methods which exploit different assumptions about the residue.

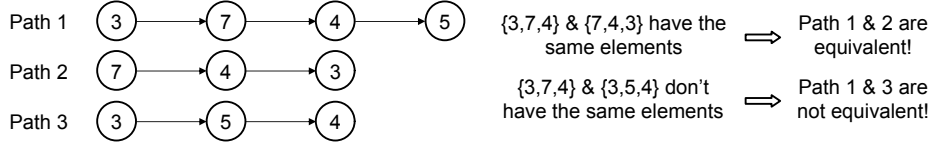


Figure 3: Path Equivalency: Path 1 and Path 2 are equivalent as first three nodes of Path 1 contain only nodes in Path 2. Path 3 is not equivalent to Path 1 as the node '5' is not an element of the first three nodes of Path 1. Note that orthogonal projections ensure Path 2 to select '5' as the next node, while there is no guarantee that Path 3 will select '7'.

#### 4.3.1. Additive Cost Model

The additive cost model assumes that the  $K$  vectors in the representation make on average equal contributions to  $\|\mathbf{y}\|_2$ . That is, we assume that the average contribution of a vector is  $\delta_e = \|\mathbf{y}\|_2 / K$ . Then, the unopened  $K - l$  nodes of a partial path of length  $l$  are expected to reduce  $\|\mathbf{r}\|_2$  by  $(K - l)\delta_e$ . Combining this with (8), the auxiliary function should satisfy

$$d_{add}(\mathbf{S}^l) \geq (K - l) \frac{\|\mathbf{y}\|_2}{K}. \quad (11)$$

Consequently, we define the additive auxiliary function as

$$d_{add}(\mathbf{S}^l) \triangleq \beta(K - l) \frac{\|\mathbf{y}\|_2}{K}, \quad (12)$$

where  $\beta$  is a constant greater than 1. Finally, we obtain the additive cost function as

$$F_{add}(\mathbf{S}^l) = \|\mathbf{r}^l\|_2 - \beta \frac{(K - l)}{K} \|\mathbf{y}\|_2. \quad (13)$$

Here,  $\beta$  acts as a regularization constant. If it is large, shorter paths are favored, making the search expand more candidates. When it becomes smaller, the search prefers longer paths. Note that favoring shorter paths increases the number of paths opened throughout the search, which improves the search at the expense of increased complexity. Hence, beta should be chosen to balance the available computational power or time restrictions and the recovery performance.

Note that  $\delta_e = \|\mathbf{y}\|_2 / K$  does not hold in general. However, (11) requires this assumption only on average. Moreover, we intuitively expect the search to miss mostly the vectors with smaller contributions to  $\|\mathbf{y}\|_2$ , and for these, the additive auxiliary function satisfies (11) with higher probabilities.

#### 4.3.2. Adaptive Cost Model

The auxiliary function can also be chosen adaptively by modifying the expectation on average contribution of an unopened node as:

$$\delta_e = (\|\mathbf{r}_i^{l-1}\|_2 - \|\mathbf{r}_i^l\|_2). \quad (14)$$

Then, the adaptive auxiliary function should fulfill

$$d_{adap}(\mathbf{S}_i^l) \geq (K - l)(\|\mathbf{r}_i^{l-1}\|_2 - \|\mathbf{r}_i^l\|_2), \quad (15)$$

where the subscript  $i$  indicates the dependency on the particular path  $\mathbf{S}_i^l$ . (15) can be justified by the fact that  $\mathbf{A}^*$  is configured to select first the vectors with higher contributions to  $\mathbf{y}$ . Hence, the residue is expected to decrease slower in later nodes than the initial nodes of a path.

As for the additive case, we incorporate  $\beta > 1$  to finally obtain the adaptive auxiliary function

$$d_{adap}(\mathbf{S}_i^l) = \beta(\|\mathbf{r}_i^{l-1}\|_2 - \|\mathbf{r}_i^l\|_2)(K - l). \quad (16)$$

The adaptive cost function can then be written as follows:

$$F_{adap}(\mathbf{S}_i^l) = \|\mathbf{r}_i^l\|_2 - \beta(\|\mathbf{r}_i^{l-1}\|_2 - \|\mathbf{r}_i^l\|_2)(K - l), \quad (17)$$

where the role of the regularization constant  $\beta$  is very similar to the additive case.



#### 4.3.3. Multiplicative Cost Model

In contrast to addition of the auxiliary function, multiplicative cost model path employs a weighting function. Here, we assume that each node reduces  $\|\mathbf{r}\|_2$  by a constant ratio,  $\alpha$ . The multiplicative cost function is defined as

$$F_{mul}(\mathbf{S}_i^l) = \alpha^{K-l} g(\mathbf{S}_i^l) = \alpha^{K-l} \|\mathbf{r}_i^l\|_2. \quad (18)$$

where  $\alpha$  should be chosen between 0 and 1. The role of  $\alpha$  is very close to that of  $\beta$  for the additive cost function. When  $\alpha$  is close to 0, short paths are assigned very small costs, making the search to prefer them. On the contrary, if we choose  $\alpha$  close to 1, weighting is hardly effective on the cost function, hence longer paths will be favored.

In contrast to the additive one, adaptive and multiplicative cost models adjust the expected decay in  $\mathbf{r}_i^l$  dynamically throughout the search. These dynamic structures are expected to provide a better modeling of the decrease in  $\mathbf{r}_i^l$ . In fact, the simulation results in Section 5 clearly indicate that they improve the reconstruction accuracy.

#### 4.4. A\* Orthogonal Matching Pursuit

We can now outline A\*OMP:  $I$  out of the  $P$  paths, which are kept in a stack, are initialized as the  $I$  vectors which best match  $\mathbf{y}$  and the remaining  $P - I$  paths are left empty. The cost for the empty paths is  $\|\mathbf{y}\|_2$ , hence they will be removed first. In each iteration, first, we select the path with minimum cost. We, then, expand the best  $B$  children of the selected path applying the pruning rules discussed in Section 4.2. Iterations are run until the selected path has length  $K$ . The pseudo-code for the algorithm is given in Algorithm 1.

We note that other termination criteria are also possible, including, for example, norm of the residue falling below a threshold, or no further reduction of the residue obtained.

#### 4.5. Complexity vs. Accuracy

The complexity of A\*OMP approach arises from two points: The number of inner-product checks between the residue and dictionary atoms, and the number of orthogonal projections. The number of inner-product checks is equal to the number of iterations. Orthogonal projection, on the other hand, is necessary for each path, except the ones that are pruned by the equivalent path pruning. Hence, the number of these is equal to  $B$  times the number of iterations minus the number of equivalent paths detected. Consequently, the important factors that govern the complexity of A\*OMP are, first, the number of iterations and, second, the number of equivalent paths detected. However, it is not possible to find reasonable approximations of these. The only approximation to the number of paths is the upper bound that assumes opening of every possible node on the tree, which is obviously far away from being realistic. In order to give an insight on these, we investigate these experimentally in section 5.1.1.

The pruning strategies of Section 4.2 can be seen as a trade-off between the accuracy and complexity of A\*OMP. If we set  $I = N$ ,  $B = N$  and  $P = \infty$ , the algorithm will perform an exhaustive search, which is prohibitively complex. On the other hand, setting  $I = 1$  and  $B = 1$  yields OMP. A choice between the accuracy and complexity of the search can be adjusted by the pruning parameters. The accuracy is expected to increase with increasing these parameters, as demonstrated in section 5.1.3. In practice, these parameters, of course, may not be increased after some point, and regarding the results in section 5.1.3, it is also questionable if they will improve the performance after some point.

The cost model is also extremely important in the complexity-accuracy trade-off. An appropriate modeling of the decay in the residue improves the ability to predict branches on which the solution might lie. Therefore, the auxiliary function is important for both choosing the best path and pruning. With an appropriate choice, the trade-off between the complexity and accuracy is boosted in favor of accuracy, such as the dynamic cost functions improving the reconstruction ability in the first example in section 5. In addition, the auxiliary function parameters  $\alpha$  and  $\beta$  also affect the complexity-accuracy trade-off. Choosing  $\beta \gg 1$  or  $0 < \alpha \ll 1$  makes the search favor shorter paths, leading to improvements in accuracy with longer search times. On the contrary, when  $\beta$  and  $\alpha$  are close to 1, the algorithm performs similar to OMP. These improvements are, of course, also expected to have some limits, for example, decreasing  $\alpha$  does not improve the performance after some point, as demonstrated in section 5.1.3.

In order to get the best out of the search parameters, they should better be considered together. For example, reducing  $\alpha$  increases the number of paths opened throughout the search. Consequently, a lower  $\alpha$  value should be accompanied by an increment in the beam width  $P$  in order to obtain better reconstruction results. This also holds when  $\beta$  or  $B$  is increased, which similarly increases the number of paths involved in the search. Examples in section 5.1.3 illustrate this issue.

---

**Algorithm 1** A\* ORTHOGONAL MATCHING PURSUIT
 

---

**Define:**

$P$  := Maximum number of search paths  
 $I$  := Number of initial search paths  
 $B$  := Number of extended branches per iteration  
 $\mathbf{S}_i = \{\mathbf{s}_i^l\}$ , matrix of atoms  $\mathbf{s}_i^l$  on the  $i$ 'th path  
 $\mathbf{c}_i = \{c_i^l\}$ , vector of coefficients for the atoms on the  $i$ 'th path  
 $L_i$  := length of the  $i$ 'th path  
 $C_i$  := cost for selecting the  $i$ 'th path

**Initialize:**

$\mathbf{T} \leftarrow \emptyset$   
**for**  $i \leftarrow 1$  to  $I$  **do** ▷  $I$  paths of length 1  
    $\hat{n} \leftarrow \arg \max_{n, \mathbf{v}_n \in \Phi \setminus \mathbf{T}} \langle \mathbf{y}, \mathbf{v}_n \rangle$   
    $\mathbf{T} \leftarrow \mathbf{T} \cup \mathbf{v}_{\hat{n}}$   
    $\mathbf{s}_i^1 \leftarrow \mathbf{v}_{\hat{n}}, c_i^1 \leftarrow \langle \mathbf{y}, \mathbf{v}_{\hat{n}} \rangle$   
    $\mathbf{r}_i \leftarrow \mathbf{y} - c_i^1 \mathbf{s}_i^1$   
    $C_i = F(\mathbf{S}_i), L_i = 1$   
**end for**  
 $C_i = \|\mathbf{y}\|_2, L_i = 0, \forall i = I + 1, I + 2, \dots, P$   
 $best\_path \leftarrow 1$   
**while**  $L_{best\_path} \neq K$  **do**  
    $\hat{p} \leftarrow best\_path$  ▷ first to replace  
    $\mathbf{T} \leftarrow \mathbf{S}_{best\_path}$   
   **for**  $i \leftarrow 1$  to  $B$  **do** ▷ extensions per path pruning  
      $\hat{n} \leftarrow \arg \max_{n, \mathbf{v}_n \in \Phi \setminus \mathbf{T}} \langle \mathbf{r}_{best\_path}, \mathbf{v}_n \rangle$   
      $\mathbf{T} \leftarrow \mathbf{T} \cup \mathbf{v}_{\hat{n}}$   
      $\hat{\mathbf{S}} \leftarrow \mathbf{S}_{best\_path} \cup \mathbf{v}_{\hat{n}}$  ▷ candidate path  
      $\hat{\mathbf{c}} \leftarrow \arg \min_{\alpha} \|\mathbf{y} - \hat{\mathbf{S}}\alpha\|_2$  ▷ Orthogonal projection  
      $\hat{C} \leftarrow F(\hat{\mathbf{S}})$  ▷ Cost of the candidate path  
     **if**  $(\hat{C} < F(\mathbf{S}_{\hat{p}})) \ \&$  ▷ tree size pruning  
      $(\hat{\mathbf{S}} \neq \mathbf{S}_j, \forall j = 1, 2, \dots, P)$  **then** ▷ path equivalency  
        $\mathbf{S}_{\hat{p}} \leftarrow \hat{\mathbf{S}}, \mathbf{c}_{\hat{p}} \leftarrow \hat{\mathbf{c}}, C_{\hat{p}} \leftarrow \hat{C}$   
        $L_{\hat{p}} \leftarrow L_{best\_path} + 1$   
        $\mathbf{r}_{\hat{p}} \leftarrow \mathbf{y} - \mathbf{S}_{\hat{p}} \mathbf{c}_{\hat{p}}$   
        $\hat{p} \leftarrow \arg \max_{i \in \{1, 2, \dots, P\}} C_i$  ▷ to be replaced next  
     **end if**  
   **end for**  
    $best\_path \leftarrow \arg \min_{i \in \{1, 2, \dots, P\}} C_i$  ▷ select best path  
**end while**  
**return**  $\mathbf{S}_{best\_path}, \mathbf{c}_{best\_path}$

---

## 5. Simulation Results

We demonstrate sparse recovery via A\*OMP in two problems in comparison to BP, SP and OMP. First of them is the recovery of a synthetically generated 1D signals, while the latter involves an image reconstruction problem. The simulations for A\*OMP were performed using the AStarOMP software developed by the authors. The AStarOMP software incorporates a trie structure to implement the A\* search tree in an efficient way. The orthogonalization over the residue is solved using the QR factorization. This software, and its MATLAB version, are available at

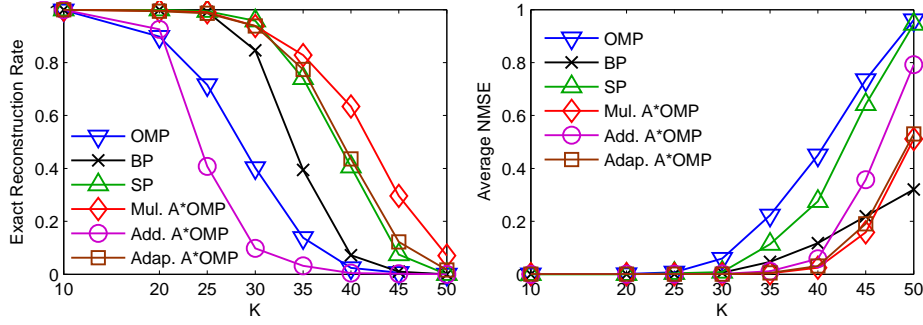


Figure 4: Reconstruction results over sparsity for uniform sparse signals employing Gaussian observation matrices.

<http://myweb.sabanciuniv.edu/karahanoglu/research/>.

### 5.1. Reconstruction of Synthetically Generated 1D Data

In this section, we evaluate three versions of A\*OMP using additive, adaptive and multiplicative cost models. These are abbreviated as Add-A\*OMP, Adap-A\*OMP and Mul-A\*OMP, respectively. The experiments cover different non-zero coefficient distributions, including uniform and Gaussian distributions as well as binary nonzero coefficients. We investigate reconstruction via Gaussian and Bernoulli observation matrices and compare different A\*OMP parameters. Finally, we demonstrate A\*OMP for reconstruction from noisy observations.

All the simulations in this section were repeated over 500 randomly generated  $K$ -sparse samples of length  $N = 256$  from which  $M = 100$  random observations were taken via the observation matrix  $\Phi$ . Reconstruction accuracy are given in terms of both the exact reconstruction rate and the average normalized mean squared error (NMSE), which is defined as the average ratio of the  $\ell_2$  norm of the reconstruction error to  $\|\mathbf{x}\|_2$  over the 500 test samples. For the noisy scenarios, we give the reconstruction error in the decibel scale, which we call the distortion ratio. Unless given explicitly, the following are common in all simulations: A\*OMP parameters were set as  $I = 3$ ,  $B = 2$ ,  $P = 200$ ,  $\beta = 1.25$  and  $\alpha = 0.8$ . For each test sample, we employed an individual observation matrix  $\Phi$  whose entries were drawn from the Gaussian distribution with mean 0 and standard deviation  $1/N$ .

#### 5.1.1. Different Coefficient Distributions

The first set of simulations employ sparse signals with nonzero coefficients drawn from the uniform distribution  $U[-1, 1]$ . We refer to these signals as uniform sparse signals in the rest. The results of these simulations for  $K$  from 10 to 50 are depicted in Fig. 4. In this test, Adap-A\*OMP and Mul-A\*OMP clearly provide lower average NMSE than BP, SP and OMP, except for  $K = 50$  where BP provides lower error. As expected, the average NMSE of OMP is the worst, while that of SP is only slightly better. BP provides lower error than SP and OMP, however it is still worse than A\*OMP except for  $K = 50$ . Even the Add-A\*OMP, which employs no dynamic cost model, yields lower error than BP up to  $K = 40$ . In addition to average NMSE, Mul-A\*OMP, on general, yields higher exact recovery rates. Though SP yields high average NMSE, its exact recovery frequency competes with that of Mul-A\*OMP up to  $K = 30$ , and even exceeds it slightly at  $K = 30$ . For Add-A\*OMP, the situation is contrary: Despite low average NMSE values, its exact reconstruction rate is even worse than that of OMP. These results indicate that the static cost model of Add-A\*OMP most of the time fails at small nonzero coefficients. Adaptive and multiplicative cost models, which dynamically adjust the expected decay in  $\|\mathbf{r}\|_2$  individually for each path, are clearly more effective for compensating path length differences.

As for SP, the exact recovery rate is much better than the NMSE suggests. This indicates that the amount of error SP makes per failure is much higher than that of the A\*OMP algorithm. To visualize this fact, the probability density estimates of the error are depicted in Fig. 5 for SP and Mul-A\*OMP. These were computed using Gaussian kernels over NMSE of the test vectors which could not be exactly reconstructed for  $K = 30$ . The figures state that NMSE values on the order of  $10^{-3}$ 's for Mul-A\*OMP, while for SP, they range up to 0.8, with mean about 0.3. This arises from the difference in the average number of misidentified elements per failure, which is shown in Fig. 6 for  $K = 30$ . Mul-A\*OMP has misidentified only one or two of the 30 components, while SP has missed 9 to 16 components, and

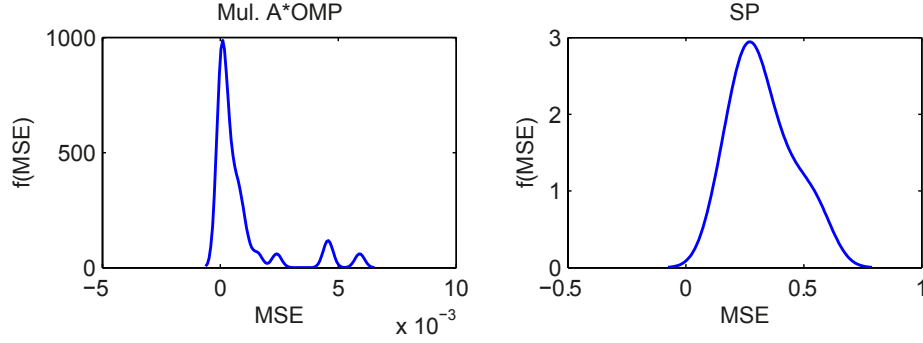


Figure 5: Probability density estimates of the NMSE for  $K = 30$ .

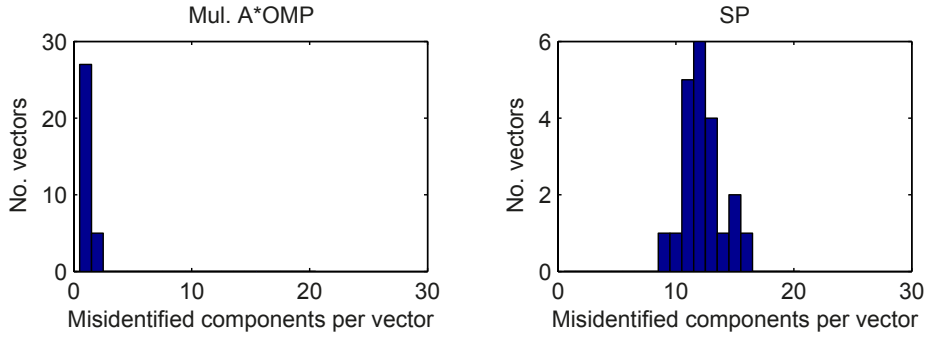


Figure 6: Number of misidentified entries per test sample for  $K = 30$ .

on average about 12 per failure. These figures indicate that if the reconstruction is not exact, SP almost completely fails, however A\*OMP can still reconstruct the desired vector with small amount of error, which is less than 1% of the signal norm for  $K = 30$ .

As discussed in section 4.5, the two important factors for the complexity of A\*OMP are the average A\*OMP iterations per vector and the average equivalent paths detected per vector. Table 1 states the average A\*OMP iterations per vector in this scenario in comparison to the upper bound on the number of A\*OMP iterations. This upper bound can easily be obtained as  $I \cdot (2^{K-1} - 1)$  for  $B = 2$  by assuming that all of the opened partial paths are selected one by one as the best path throughout the search. The actual number of iterations is incomparably lower than this upper bound. Moreover, though the upper bound increases exponentially with  $K$ , the actual number of iterations exhibit a much lower slope. The second important factor, the average number of equivalent paths per vector is given in Table 2. These numbers are comparable to the number of iterations, which states the effectiveness of the equivalent path pruning rule. These results indicate that pruning and proper selection of the cost model make it possible to run the search for cases where the upper bound becomes unpractically high.

Table 1: Average A\*OMP iterations per vector for uniform sparse signals

	$K$			
	10	20	30	40
Mul-A*OMP	13.8	164	1695	4177
Adap-A*OMP	19	167.4	2443	6109
Upper Bound	1533	$1.57 \cdot 10^6$	$1.61 \cdot 10^9$	$1.65 \cdot 10^{12}$

Finally, in order to provide an insight about the speed of the search, we list in Table 3 the average run-times for Mul-A\*OMP, Adap-A\*OMP and OMP on a modest Pentium Dual-Core CPU at 2.3GHz. These were obtained

Table 2: Average equivalent paths per vector for uniform sparse signals

	$K$			
	10	20	30	40
Mul-A*OMP	4.4	114.1	975.2	1776
Adap-A*OMP	11.2	126.6	1355	1831

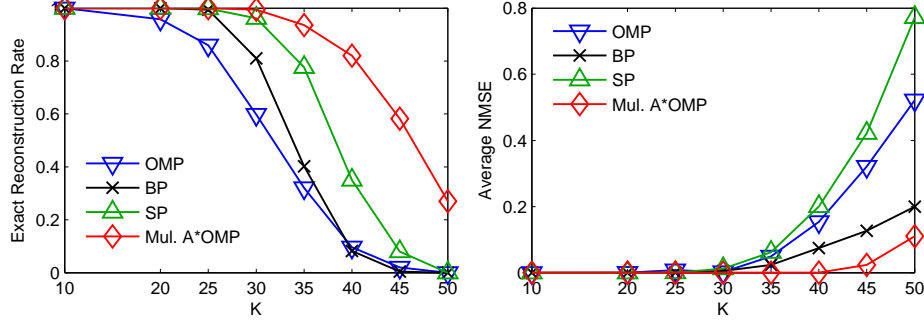


Figure 7: Reconstruction results over sparsity for Gaussian sparse vectors using Gaussian observation matrices.

using the AStarOMP software and a similar OMP implementation developed by the authors specially for obtaining comparable run-times. Note that the structure of A\*OMP makes it possible to process the  $B$  candidates in parallel at each iteration. Moreover, the search can easily be modified to open more than one promising path per iteration in parallel. Hence, these run-times can be significantly reduced by parallel programming, which is beyond the scope of this paper.

Table 3: Average run-time in sec. per vector for uniform sparse signals

	$K$			
	10	20	30	40
OMP	0.0012	0.0025	0.0036	0.0050
Mul-A*OMP	0.0022	0.0261	0.3158	0.8292
Adap-A*OMP	0.0032	0.0276	0.4601	1.1525

For the second set of simulations, we employ Gaussian sparse vectors, whose nonzero entries were drawn from the standard Gaussian distribution. Fig. 7 depicts the average NMSE and exact reconstruction rates for this test. In this scenario, Mul-A\*OMP provides clearly better reconstruction than BP, SP and OMP. We observe that it provides both lower NMSE and higher exact reconstruction rate than all the other algorithms. SP yields the second best exact reconstruction rate, however, its average NMSE is the worst, as a consequence of the almost complete failure of a non-exact reconstruction.

In order to question the choice of the observation matrix, we repeat the last scenario with observation matrices drawn from the Bernoulli distribution. The average NMSE and exact reconstruction rates for this test are illustrated in Fig. 8. Comparing Fig. 8 with Fig. 7, we observe that the average NMSE values remain quite unaltered for Mul-A\*OMP and BP, while that for SP increases. Mul-A\*OMP leads to the least amount of error. As for exact reconstruction, only BP keeps the same rates, while the rates of all others fall. BP and SP compete with Mul-A\*OMP until  $K = 25$ , where SP is slightly better. When  $K$  further increases, Mul-A\*OMP has the highest exact recovery frequency.

Next problem is the reconstruction of sparse binary vectors, where the nonzero coefficients were selected as 1. The results are shown in Fig. 9. We observe that BP clearly yields better reconstruction than the others in this case. SP also performs better than A\*OMP. The failure of A\*OMP is related to the fact that this is a particularly challenging case for OMP-type of algorithms [6]. OMP is shown to have non-uniform guarantees, and, though mathematical justification of A\*OMP is quite hard, this non-uniformity seems to be carried over to A\*OMP for this type of signals.

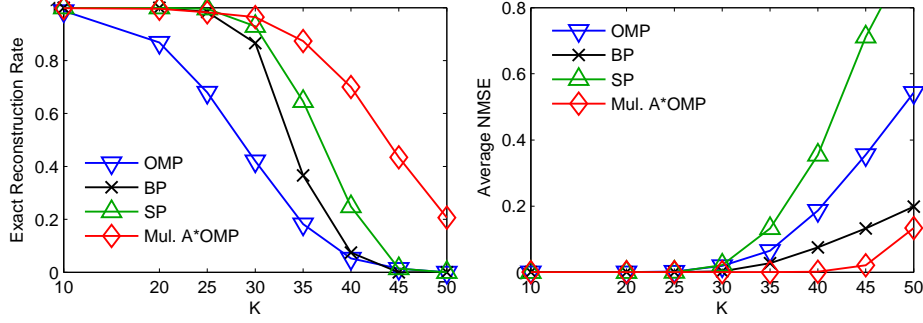


Figure 8: Reconstruction results over sparsity for Gaussian sparse vectors using Bernoulli observation matrices.

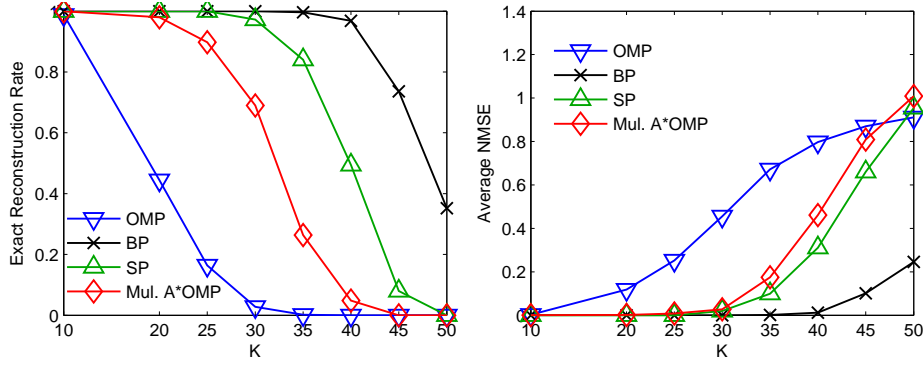


Figure 9: Reconstruction results over sparsity for sparse binary signals using Gaussian observation matrices.

In contrast, for sparse binary signals,  $\ell_0$  norm of the correct solution is exactly equal to its  $\ell_1$  norm, which might be considered as an advantage for BP in this particular scenario. The results of this scenario, however, should not be very discouraging since sparse binary vectors represent a limited subset of the real world problems.

### 5.1.2. Performance over Different Observation Lengths

Another interesting test case is the reconstruction ability when the observation length,  $M$ , changes. Fig. 10 depicts the recovery performance over  $M$  for uniform sparse signals where  $K = 25$ . For each  $M$  value, a single Gaussian observation matrix is employed to obtain observations from all signals. We observe that Mul-A\*OMP is the best in terms of the exact recovery rates, while SP and BP compete it for  $M \geq 90$  and  $M \geq 100$ , respectively. The average NMSE of Mul-A\*OMP is also lower than the others except for the case of  $M = 50$  where BP provides lower error than Mul-A\*OMP.

### 5.1.3. Comparison of Different Search Parameters

Choosing the search parameters is an important issue for the A\*OMP algorithm. This was discussed above in section 4.5, indicating two main points: The reconstruction performance of the search might be increased by modifying the search parameters to explore more paths in the search at the expense of increased iterations and search times. In order to demonstrate this, we consider two scenarios. First, we vary  $\alpha$ , and later  $B$  together with  $P$ .

Fig. 11 depicts the performance of Mul-A\*OMP over  $\alpha$  for uniform sparse signals with  $K = 30$  and  $K = 35$ . The dashed and solid lines indicate results for  $P = 200$  and  $P = 5000$ , respectively. For  $K = 30$ , the reconstruction performance increases when  $\alpha$  is reduced from 0.95 to about 0.8, whereas any further reduction of  $\alpha$  does not significantly affect the performance. In addition, there is hardly any difference between selecting  $P = 200$  and  $P = 5000$ . This suggests that setting  $P = 200$  and  $\alpha \approx 0.8$  seems to be enough for  $K = 30$ . When  $K = 35$ , however, more paths are involved in the search, and increasing  $P$  improves the reconstruction. When  $P = 200$ , reducing  $\alpha$  below 0.9 does not improve but slightly degrade the performance. On the contrary, if  $P$  is increased to 5000, the reconstruction is

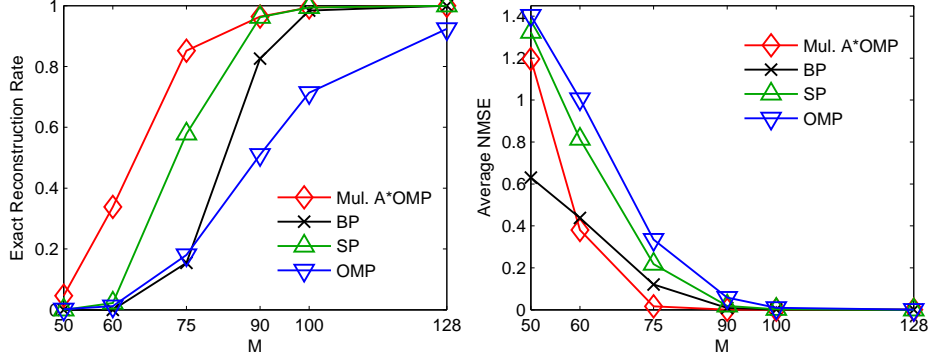


Figure 10: Reconstruction results over observation lengths for uniform sparse signals where  $K = 25$  using a single Gaussian observation matrix for each  $M$ .

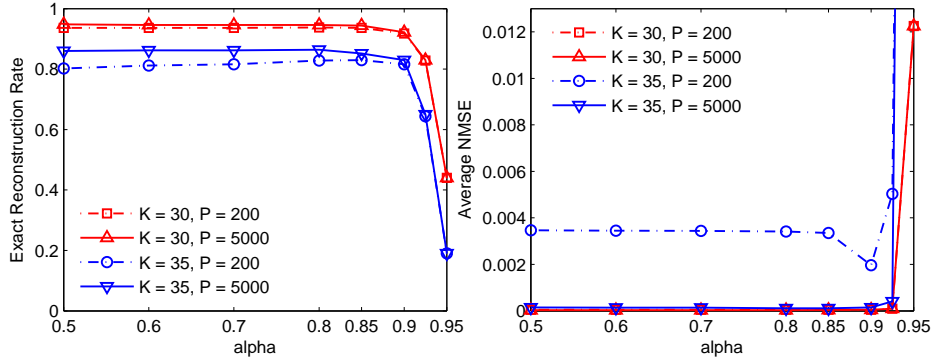


Figure 11: Reconstruction results over  $\alpha$  for uniform sparse signals using Gaussian observation matrices.

Table 4: Average Mul-A\*OMP iterations per vector wrt.  $\alpha$  and  $P$  for uniform sparse signals with  $K = 35$

	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.7$	$\alpha = 0.8$	$\alpha = 0.9$
$P = 200$	4158	3927	3565	2932	1353
$P = 5000$	58204	51710	41781	25527	4026

improved until  $\alpha$  is reduced to 0.8, below which the reconstruction performance does not change. Though not given in the figures, the authors have observed that setting  $P > 5000$  has hardly any effect on the reconstruction. These results demonstrate that reducing  $\alpha$  improves the reconstruction until some convergence point. Table 4 lists the average number of search iterations while  $\alpha$  and  $P$  are varied. We observe that decreasing  $P$  and increasing  $\alpha$  increase the number of paths involved, which clarifies complexity-accuracy trade off that leads to improved recovery performance at the expense of increased complexity.

Next, we illustrate the performance of Mul-A\*OMP with  $B = 2$  and  $B = 3$  for sparse binary signals in Fig. 12. The experiment was repeated for  $P = 200$  and  $P = 1000$ , which are depicted by dashed and solid lines, respectively. We observe that increasing  $B$  from 2 to 3 improves the reconstruction. This improvement is further enhanced by increasing  $P$  from 200 to 1000 when  $K \geq 25$ , where a larger search stack can better cover for the increased number of paths involved in the search. Table 5 lists the average number of search iterations, which increase with  $B$  and  $P$ . Hence, the improvement, as above, is obtained at the expense of complexity.

The results in this section explain how the performance of A\*OMP can be adjusted by the search parameters. The mechanism behind is simple: Increasing the number of paths explored by the search improves the results, until a convergence point, at the expense of increasing the complexity. According to the experimental results, one advantage is that even with modest settings such as  $I = 2$ ,  $P = 200$  and  $\alpha = 0.8$  employed in the experiments, A\*OMP can

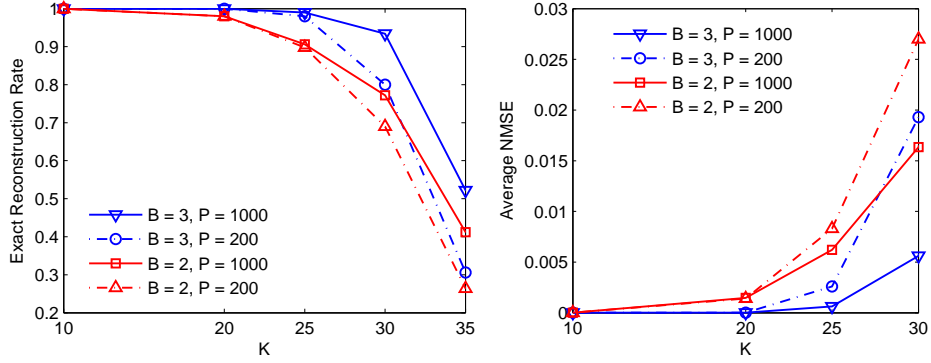


Figure 12: Reconstruction results for sparse binary signals for  $B = 2$  and  $B = 3$  using Gaussian observation matrices.

Table 5: Average Mul-A\*OMP iterations wrt.  $B$  and  $P$  per vector in the sparse binary problem

	P = 200		P = 1000	
	B=2	B=3	B=2	B=3
$K = 10$	48	114	48	114
$K = 20$	1046	2095	1275	7159
$K = 30$	3424	4249	12278	18240

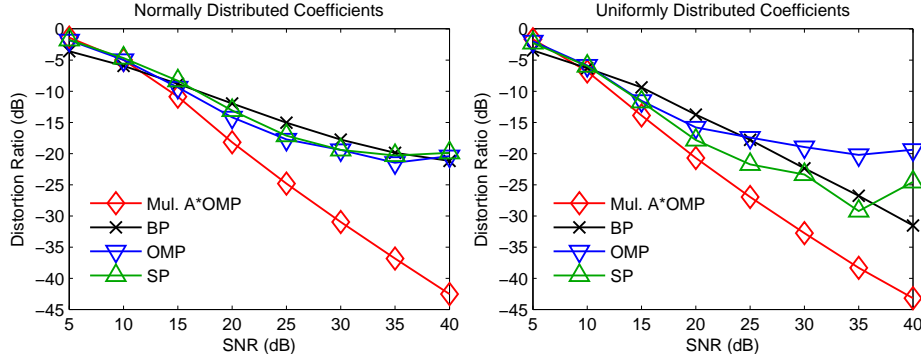


Figure 13: Average NMSE over SNR for reconstruction of sparse signals from noisy observations using Gaussian observation matrices.

provide high exact recovery frequencies and lower error than the other candidates for uniform and Gaussian sparse signals. This indicated that A\*OMP recovery, at least in these cases, is quite robust against the choice of search parameters.

#### 5.1.4. Reconstruction from Noisy Observations

Fig. 13 illustrates recovery results where the observation vectors are contaminated by white gaussian noise at different SNR levels. Here,  $K$  is 25 and 30 for Gaussian and uniform sparse signals, respectively. The results are shown in terms of the distortion ratio in the decibel scale for better comparison. We observe that Mul. A\*OMP produces less error than BP, SP and OMP for about 10dB and higher. When SNR decreases, BP starts to be more effective than the greedy algorithms.

#### 5.2. Reconstruction of Images

We finally simulate the reconstruction ability of A\*OMP on some commonly used  $512 \times 512$  images including 'Lena', 'Tracy', 'cameraman', etc. The images were reconstructed in  $8 \times 8$  blocks which provide important advantages that reduce the complexity and memory requirements of the search. First, without block-processing, the reconstruction



Table 6: PSNR values for images reconstructed using different algorithms

	BP	OMP	SP	Mul-A*OMP		Adap-A*OMP	
				B=2	B=3	B=2	B=3
Lena	33.5	29.6	27.5	36.4	38.3	35.2	37
Tracy	40.6	36.8	33.9	44.8	46.4	44.5	45.5
Pirate	31.7	27.7	25.3	33.6	34.5	32.8	34.2
Cameraman	34.4	30.7	28.5	38.4	40.2	36.7	39.5
Mandrill	28.3	24.4	22.1	30.3	31.3	29.3	30.8

problem requires searching among  $N = 512^2 = 262144$  dictionary atoms. However, block-processing reduces the problem to 4096 subproblems with  $N = 64$ , which is more efficient as each subproblem requires a search in 4096-fold reduced dimensionality. Second, block-processing reduces the total number of search paths drastically. To illustrate, let's set  $B = 2$ . From Section 4.2, the number of search paths for each  $K$ -sparse block is upper bounded by  $I \cdot 2^{(K-1)}$ . Then, for the whole image, the upper bound becomes  $4096 \cdot I \cdot 2^{(K-1)} = I \cdot 2^{(K+1)}$ . If no block processing were involved, the upper bound would be  $I \cdot 2^D$  where  $D \gg K + 1$ . Finally, block-processing also reduces the length of the involved paths. Note that the block structure is shared by all involved recovery methods.

The simulations were performed with five  $512 \times 512$  grayscale images using the 2D Haar Wavelet basis  $\Psi$ . Note that in this case, the dictionary is not  $\Phi$ , but the holographic basis  $\mathbf{V} = \Phi\Psi$ . Images were first preprocessed such that each  $8 \times 8$  block is  $K$ -sparse in the 2D Haar Wavelet basis, where  $K = 14$ . A single observation matrix  $\Phi$  of size  $M \times N$ , which was randomly drawn from the Gaussian distribution with mean 0 and standard deviation  $1/N$ , was employed to compute the measurements of length  $M = 32$  from each block. Mul-A\*OMP and Adap-A\*OMP were run for both  $B = 2$  and  $B = 3$ . We selected  $I = 3$  and  $P = 200$ . The cost function parameters were set to  $\alpha = 0.5$  and  $\beta = 1.25$ .

Table 6 lists the peak Signal-to-Noise ratio (PSNR) of reconstructed images. A\*OMP yields better reconstruction than the other methods. Increasing  $B$  from 2 to 3 further improves the reconstruction performance. A\*OMP improves PSNR up to 5.8 dB, and 4.4 dB on average over BP. As an example, Fig. 14 depicts reconstruction of 'lena' using SP, BP and Mul-A\*OMP with  $B = 3$ . Mul-A\*OMP reconstruction provides lower error, which can be observed better in Fig. 15 illustrating the absolute error per pixel for BP and Adap-A\*OMP reconstructions. For BP, errors are concentrated around boundaries and detailed regions, while Mul-A\*OMP clearly produces less distortion all around the image.

## 6. Conclusion

This work introduces a novel CS reconstruction approach, A\*OMP, which is based on an effective combination of OMP with A\* search. This semi-greedy method performs a tree-search, that favors the paths minimizing the cost function on-the-fly. In order to compare paths with different lengths, novel dynamic cost functions, which show better reconstruction in the provided experiments, are defined. Pruning strategies are introduced to limit the search running times. A complexity-accuracy trade-off is provided via adjustment of the search parameters. In the provided experiments, A\*OMP, with some modest settings, performs better reconstruction for uniform and Gaussian sparse signals, and for images than BP and SP. It also shows robust performance under presence of noise. BP and SP perform better than A\*OMP for the sparse binary signals which constitute a limited subset of the real world problems. Moreover, as demonstrated, the A\*OMP reconstruction in this case can be improved by modifying the search parameters at the expense of complexity.

To conclude, the demonstrated reconstruction performance of A\*OMP indicates that it is a promising approach, that is capable of reducing the reconstruction errors significantly.

## References

- [1] E. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inf. Theory 52 (2006) 489–509.
- [2] D. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (2006) 1289–1306.



Figure 14: Reconstructions of image 'Lena' using different algorithms

- [3] E. Candès, T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies?, *IEEE Trans. Inf. Theory* 52 (2006) 5406–5425.
- [4] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, *SIAM J. on Sci. Comp.* 20 (1998) 33–61.
- [5] Y. C. Pati, R. Rezaiifar, P. S. Krishnaprasad, Orthogonal matching pursuit: Recursive bfunction approximation with applications to wavelet decomposition, in: *Proc. 27th Asilomar Conference on Signals, Systems and Computers*, volume 1, Los Alamitos, CA, pp. 40–44.
- [6] W. Dai, O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, *IEEE Trans. Inf. Theory* 55 (2009) 2230–2249.

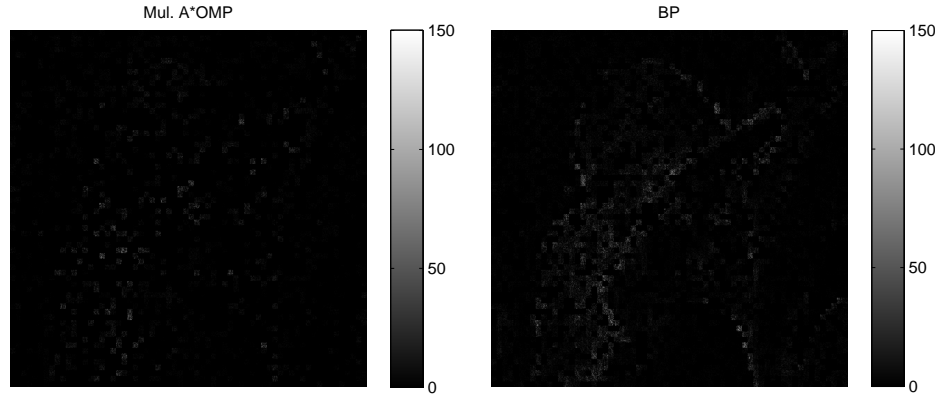


Figure 15: Reconstruction error per pixel of image 'Lena' for Mul-A\*OMP with  $B = 3$  and BP.

- [7] T. Blumensath, M. Davies, Iterative thresholding for sparse approximations, *J. Fourier Anal. Appl.* 14 (2008) 629–654.
- [8] T. Blumensath, M. Davies, Iterative hard thresholding for compressed sensing, *Appl. Comp. Harmonic Anal.* 27 (2009) 265–274.
- [9] S. Koenig, M. Likhachev, Y. Liu, D. Furcy, Incremental heuristic search in ai, *AI Mag.* 25 (2004) 99–112.
- [10] R. Dechter, J. Pearl, Generalized best-first search strategies and the optimality of A\*, *J. ACM* 32 (1985) 505–536.
- [11] F. Jelinek, *Statistical Methods For Speech Recognition*, MIT Press, Cambridge, MA, USA.
- [12] P. E. Hart, N. J. Nilsson, B. Raphael, A formal basis for the heuristic determination of minimum cost paths, *IEEE Trans. Syst. Sci. Cybern.* 4 (1968) 100–107.
- [13] P. E. Hart, N. J. Nilsson, B. Raphael, Correction to a formal basis for the heuristic determination of minimum cost paths, *SIGART Newsletter* (1972) 28–29.
- [14] N. B. Karahanoglu, H. Erdogan, Compressed sensing signal recovery via a\* orthogonal matching pursuit, in: *Acoustics, Speech and Signal Processing (ICASSP)*, 2011 IEEE International Conference on, pp. 3732–3735.
- [15] S. Cotter, B. Rao, Application of tree-based searches to matching pursuit, in: *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, pp. 3933–3936.
- [16] G. Karabulut, L. Moura, D. Panario, A. Yongacoglu, Integrating flexible tree searches to orthogonal matching pursuit algorithm, *Vision, Image and Signal Processing, IEE Proceedings - 153* (2006) 538–548.
- [17] P. Schniter, L. Potter, J. Ziniel, Fast bayesian matching pursuit, in: *Information Theory and Applications Workshop, 2008*, pp. 326–333.
- [18] M. Elad, I. Yavneh, A plurality of sparse representations is better than the sparsest one alone, *Information Theory, IEEE Transactions on* 55 (2009) 4701–4714.
- [19] H. Pham, W. Dai, O. Milenkovic, Sublinear compressive sensing reconstruction via belief propagation decoding, in: *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, pp. 674–678.
- [20] C. La, M. Do, Tree-based orthogonal matching pursuit algorithm for signal reconstruction, in: *Image Processing, 2006 IEEE International Conference on*, pp. 1277–1280.
- [21] E. Candès, M. Rudelson, T. Tao, R. Vershynin, Error correction via linear programming, in: *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, Pittsburgh, PA, pp. 295–308.
- [22] E. Candès, T. Tao, Decoding by linear programming, *IEEE Trans. Inf. Theory* 51 (2005) 4203–4215.
- [23] E. Candès, The restricted isometry property and its implications for compressed sensing, *Comptes Rendus Mathématique* 346 (2008) 589–592.
- [24] M. Rudelson, R. Vershynin, Sparse reconstruction by convex relaxation: Fourier and gaussian measurements, in: *40th Annual Conference on Information Sciences and Systems*, pp. 207–212.
- [25] J. A. Tropp, S. J. Wright, Computational methods for sparse solution of linear inverse problems, *Proc. IEEE* 98 (2010) 948–958.
- [26] S. Foucart, M.-J. Lai, Sparsest solutions of underdetermined linear systems via  $\ell_q$ -minimization for  $0 < q \leq 1$ , *Applied and Computational Harmonic Analysis* 26 (2009) 395–407.
- [27] S. Mallat, Z. Zhang, Matching pursuit in a time-frequency dictionary, *IEEE Trans. Signal Process.* 41 (1993) 3397–3415.
- [28] J. A. Tropp, A. C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Trans. Inf. Theory* 53 (2007) 4655–4666.
- [29] D. Needell, R. Vershynin, Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit, *Found Comput Math* 9 (2009) 317–334.
- [30] H. Rauhut, On the impossibility of uniform sparse reconstruction using greedy methods, *Sampl. Theory Signal Image Process.* 7 (2008) 197–215.
- [31] D. L. Donoho, Y. Tsaig, I. Drori, J.-L. Starck, Sparse solution of underdetermined linear equations by stagewise Orthogonal Matching Pursuit (StOMP), Technical Report, Statistics Dept., Stanford Univ., 2006.
- [32] D. Needell, J. A. Tropp, CoSaMP: Iterative signal recovery from incomplete and inaccurate samples, *Appl. Comp. Harmonic Anal.* 26 (2008) 301–321.
- [33] V. Cevher, S. Jafarpour, Fast hard thresholding with Nesterov's gradient method, in: *Neuronal Information Processing Systems, Workshop*

on Practical Applications of Sparse Modeling, Whistler, Canada.

[34] Y. Nesterov, Smooth minimization of non-smooth functions, *Mathematical Programming* 103 (2005) 127–152.

**Nazim Burak Karahanoglu** received his B.S. degree in Electrical and Electronic Engineering from METU, Ankara in 2003 and M.S. degree in Computational Engineering from the Friedrich-Alexander University of Erlangen-Nuremberg, Germany in 2006. He has been with the Information Technologies Institute of the Scientific and Technological Research Council of Turkey (TUBITAK) since 2008. He is also a Ph.D. student at the Electronics Engineering Department of Sabanci University, Turkey. His research interests include compressed sensing and sonar signal processing.

**Hakan Erdogan** is an assistant professor at Sabanci University in Istanbul, Turkey. He received his B.S. degree in Electrical Engineering and Mathematics in 1993 from METU, Ankara and his M.S. and Ph.D. degrees in Electrical Engineering: Systems from the University of Michigan, Ann Arbor in 1995 and 1999 respectively. He was with the Human Language Technologies group at IBM T.J. Watson Research Center, NY between 1999 and 2002. He has been with Sabanci University since 2002. His research interests are in developing and applying probabilistic methods and algorithms for multimedia information extraction.